DOCUMENT RESUME

ED 067 396 TM 001 791

AUTHOR	Kaufman, Alan S.; And Others
TITLE	Test Service Bulletin.
INSTITUTION	Psychological Corp., New York, N.Y.
REPORT NO	Bull-59
PUB DATE	May 72
NOTE	12p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Bulletins; *Correlation; *Literature Reviews; Predictive Validity; *Scoring; *Statistics; *Tests
IDENTIFIERS	*Restriction of Range

ABSTRACT

In this bulletin are an article, "Restriction of Range: Questions and Answers" BY Alan S. Kaufman and book reviews. Answers to the following questions are given in the article: What exactly is meant by "restriction of range"?; What effect does it have?; Why are correlation coefficients lower for restricted than for widespread groups?; How important in a practical sense is the fact that a test's validity coefficients may be lower for groups with wide variation on the test?; What can be done to correct for restriction of range?; What can an organization do to determine the effectiveness of a predictor if it cannot carry out a lengthy study?; and Can the formula be used to correct for low correlations whenever restriction of range occurs? (DB)

ED 067396

Test Service Bulletin

THE PSYCHOLOGICAL CORPORATION
ROGER T. LENNON, *President*

May, 1972
No. 59

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

RESTRICTION OF RANGE: QUESTIONS AND ANSWERS

ALAN S. KAUFMAN

IF somewhere a catalog were kept of instances in which the results of testing seemed disappointing, it would certainly include an entry similar to the following Sample Case: *The personnel officer of a large metals manufacturing plant wishes to evaluate the test he has been using to select electrical maintenance workers. To fill the available jobs, he has hired those applicants who scored in the upper third of the possible range of scores on the test, paying little or no attention to other considerations. Now the personnel officer wants to know whether the test is effective. He obtains performance ratings for the 40 men he hired, as soon as they have completed six months on the job. He is disappointed to find a correlation of .21 between test scores and performance ratings.*

One possible reason for this relatively low correlation is the statistical phenomenon, *restriction of range*. Yet "restriction of range," when given as an explanation for a practical problem, is often not well understood. The purpose of this discussion is to clarify the concept of restricted range, particularly as it relates to situations such as the one described above. If the personnel officer in the Sample Case had been told that his results may have been partly due to restriction of range, the following are some of the questions he might have asked.

Q. What exactly is meant by "restriction of range"?

A. "Restriction of range," as applied to test scores, is a general term which means that the test scores for a particular group are concentrated in a small portion of the possible range of scores. In statistical language, groups restricted in range have smaller standard deviations than those that are not restricted.

For example, suppose that scores on a test can vary from 0 to 47. If most members of a particular group obtained scores below 15 or 20, then the scores for that group would be restricted in range. However, a group need not perform *poorly* for their scores to be termed restricted in range; if nearly all group members achieved *high* scores, or if their scores were heavily concentrated in the *middle* of the possible score distribution, then they too would be characterized by restriction in the range of their scores. Figure 1 illustrates restriction of range and presents a discussion of the phenomenon in relation to the Sample Case.

Restricted range should not be thought of as applying only to test scores. For example, a group of five-year-olds is restricted in *age* range, while a professional basketball team is restricted in range with respect to *height* (a six-foot-tall player is ordinarily referred to as "small"). In practical settings, a foreman giving performance ratings to lathe operators or a teacher giving grades in a history course may rate all the workers or all the students just about the same, whether these are all high or all low or all about average. Similarly, a number of industrial studies have shown that in some shops the workers organize informally to make sure that no one turns out much more

TM 001 791

The contents of this Bulletin are not copyrighted; the articles (except for Figure 2 on page 3) may be quoted or reprinted without formality other than the customary acknowledgement of the Test Service Bulletin of THE PSYCHOLOGICAL CORPORATION *as the source. There is no charge for the Test Service Bulletin.*

Published from time to time by the Test Division of THE PSYCHOLOGICAL CORPORATION *in the interest of promoting greater understanding of the principles and techniques of mental measurement and its applications in guidance, personnel work, and clinical psychology, and for announcing new publications of interest. Address communications to*

304 East 45th Street
New York, N.Y. 10017

Senior Professional Staff of the Test Division

ALEXANDER G. WESMAN
Director

JEROME E. DOPPELT
Associate Director

JAMES H. RICKS, JR.
Associate Director

RICHARD G. BUCHANAN
Assistant Director

DAVID O. HERMAN
Assistant Director

DAVID H. HALL
Senior Research Associate

ESTHER R. HOLLIS
Manager, Advisory Service

ALAN S. KAUFMAN
Senior Research Associate

MYRON M. THOMASON
Consultant, Advisory Service

FAY B. KRAWCHICK
Senior Editor

work or much less work than the average, making life miserable for (or even forcing out) those who are considered "rate-busters." If the ratings or grades or production records are all about the same, then the distributions of these data are restricted in range, rather than as widespread as they could normally be.

Q. O.K., I understand what is meant by restriction of range. But what effect does it have?

A. To understand the importance of restricted range, one must remember why measurements are made in the first place. A useful test helps us to see and size up the differences among people in whatever the test may be measuring. But suppose there are no differences. Then the test will not help and using it is a waste of time.

Of course, one rarely finds a group among whom there are *literally no differences* in the ability or trait being measured. However, when the differences among individuals are relatively small, such that the distribution of test scores is restricted, the test may not be very effective. Thus, a mathematics test covering topics taught in high school would not be very useful for a group of engineering graduate students since most would obtain nearly perfect scores on the test.

Raw Score	Distribution A	Distribution B
45-47	11	11
42-44	111	111
39-41	𝍸 𝍸	𝍸 𝍸
36-38	𝍸 11	𝍸 11
33-35	𝍸 1	𝍸 1
30-32	𝍸 𝍸 11	𝍸 𝍸 11
27-29	𝍸 1111	
24-26	𝍸 𝍸 111	
21-23	𝍸 𝍸 11	
18-20	𝍸 111	
15-17	𝍸 1111	
12-14	𝍸 11	
9-11	𝍸 1111	
6-8	1111	
3-5	𝍸 1	
0-2	111	
	Total Group of Applicants	*Total Group of Accepted Applicants*
	N = 120	N = 40
	Mean = 23.6	Mean = 36.4
	SD = 11.4	SD = 4.6

FIGURE 1. Hypothetical distributions on a 47-item test, illustrating the phenomenon of range restriction. If this test had been used in the Sample Case, Distribution A might represent the scores obtained by the total group of *applicants* for the job. The horizontal line shown in Distribution A separates the upper one-third of the scores from the lower two-thirds. Since the upper third of applicants were hired, the individuals whose scores are above the line represent the group of *accepted applicants*. Distribution B shows this group of accepted applicants separately. Clearly, the 40 scores in Distribution B cluster much more closely than the 120 scores in Distribution A. Only a small (or narrow or "restricted") part of the range of ability on the test is represented by the selected individuals. That is, the test scores obtained by the accepted applicants are *restricted in range*.

One way in which the test's relative lack of effectiveness will reveal itself is in its correlations with other measures. *When a group's scores on a test are restricted in range, correlations between the test and other measures will be lower for that group than for groups whose scores are not restricted.*

The correlation coefficients that are used to express reliability and validity may be low for a test administered to a homogenous group, as a direct result of the restriction of range. Since test users are often interested in the

test's effectiveness in predicting a criterion, restriction of range can present a problem.

Q. Why are correlation coefficients lower for restricted than for widespread groups?

A. The lower correlations for restricted groups can best be understood by examining the nature of the correlation coefficient itself. Figure 2 illustrates correlations of varying magnitudes between two measures. Each diagram consists of numerous points plotted on a pair of axes. A point stands for a person who has been measured on each of the variables being correlated, e.g., test and criterion; the position of the point indicates the person's test score and his performance rating.

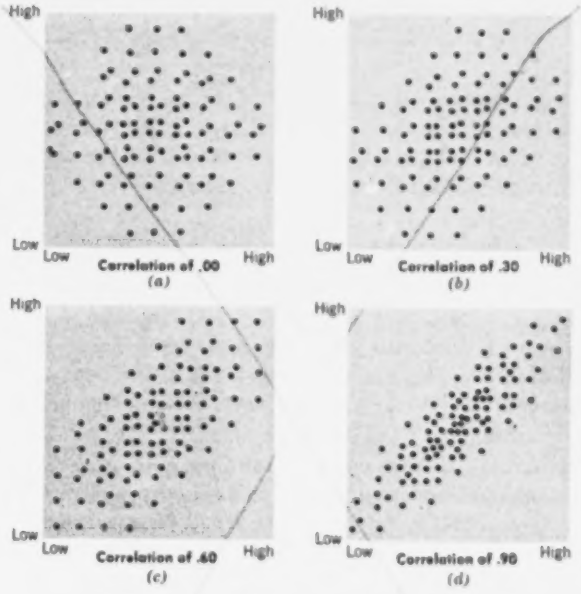


FIGURE 2. Illustration of four different levels of the correlation coefficient. In *(a)*, a correlation of .00 is shown, indicating no relationship between the variables represented by the horizontal and vertical axes. The points assume a shape that is approximately round. (Each point represents an individual's score on the two measures being correlated.) In *(b)*, a correlation of .30 is shown, indicating a moderate relationship between the variables; the points approximate the shape of a wide ellipse. Correlations of .60 and .90 are shown in *(c)* and *(d)*, respectively. Clearly, the greater the relationship between two variables, the narrower the ellipse formed by the points. In the extreme case of a perfect correlation (1.00), not shown here, the points would fall on a straight line. (From MEASUREMENT AND EVALUATION IN PSYCHOLOGY AND EDUCATION, Third Edition, by R. L. Thorndike and E. Hagen. Copyright © 1955, 1961, 1969 by John Wiley & Sons, Inc. Reprinted by permission.)

For example, suppose test scores are indicated on the horizontal axis, and performance ratings on the vertical axis of each diagram in Figure 2. Then, a point in the upper right-hand corner of a diagram would indicate a person with a high test score and a high performance rating, while a point in the lower right-hand corner would signify an individual with a high score and a low rating.

For a perfect correlation of 1.00 (not shown here), the points would form a straight line going from the lower left to the upper right of the diagram. Such a correlation, virtually never realized in practice, means that one variable may be used to predict perfectly the other variable (i.e., knowing the test score, we could predict the performance rating with perfect accuracy). As can be seen in Figure 2, the plotted points for correlations of .30, .60, and .90 form progressively narrower ellipses; for a .00 correlation, which means that the two variables are completely unrelated, the points tend to form a circle or to fill in the space evenly. In cases where two variables are related, the "skinnier" or more "line-like" the ellipse, the greater the relationship. As the ellipse approaches a straight line, there is an increasing tendency for individuals at a certain level (e.g., good, average, poor) on one of the measures to be at the same level on the second measure.

Figure 3 shows what can happen when a group is restricted in range on one of the variables being correlated. In this figure, a substantial relationship is shown between the predictor (test score) and the criterion (performance rating) for individuals represented by points covering the full range of both variables.

Also shown are the dramatic changes that occur in the pattern of the plotted points when only a restricted portion of the sample is considered. The relatively thin ellipse that is pictured for the complete group—indicative of a high correlation—disappears after the partitioning. What remain for groups restricted in range, regardless of ability level, are shapes that are closer to having about equal length and width, indicative of low correlations. Thus, correlation coefficients that are high for non-restricted groups may diminish substantially when only a portion of the whole group is considered.

If restriction occurs on the *criterion*—e.g., if a supervisor gives high performance ratings to all employees—the effect on the correlation coefficient is similar to that

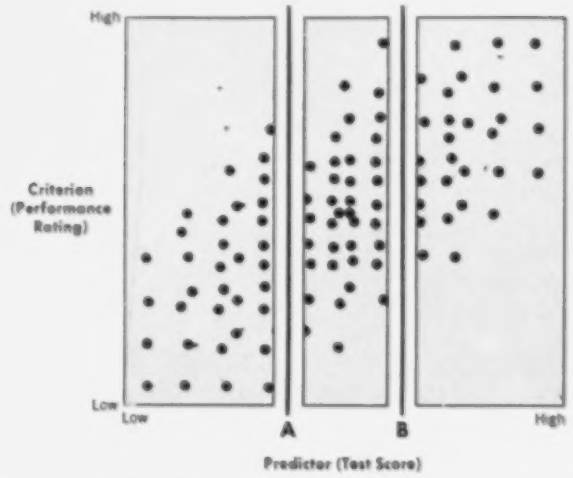


FIGURE 3. The effects of restriction of range of predictor test scores on the validity coefficient. When the range is not restricted —when scores are found in most or all of the possible categories from low to high—the diagram may be considered in its entirety and shows a correlation. This is evidenced by the elliptical shape formed by the plotted points. (These points are the same as the ones in *(c)* of Figure 2, showing a .60 correlation.)

However, the correlation diminishes greatly when only a limited group of individuals is considered, such as those scoring low on the predictor test (represented by the points to the left of line A), those scoring high on the test (represented by the points to the right of line B), or those scoring at an average level (the points between lines A and B). For these groups, restricted in the range of test scores, the points assume shapes that are not at all elliptical, indicating lower correlations between the predictor and criterion.

caused by restriction in the range of test scores. If restriction occurs on both variables, as is frequently the case in practical settings, then the decrease in the correlation tends to be even more pronounced. Accordingly, it is possible for a test that shows low validity in a restricted group to be very effective in a more heterogenous population.

One should not infer, however, that substantial correlations usually or always exist between predictors and criteria for heterogenous samples. Some would-be predictors are inefficient or inappropriate for certain purposes, and would correlate about zero with the criteria even if the test scores and the ratings were widely spread out. For example, a test of mental ability would probably be a poor predictor of performance at an unskilled job requiring simple motor coordination and little thinking.

Q. I can see that a test's validity coefficients may be lower for groups restricted in range than for groups with wide variation on the test. But how is this important in a practical sense?

A. The effects of restriction of range become particularly important from a practical standpoint where selection is made on the basis of test scores. The Sample Case, where the personnel officer hired the upper one-third of the applicants, presents such a situation. Let us consider Figure 3 as representing a similar case, with all of the points on the entire graph representing a group of *applicants*. Then, the selected group of applicants—those actually hired—might be represented by the points to the right of line B. The points to the left of the line would then represent the applicants who were not accepted for jobs at the company. Naturally, no ratings on the criterion could be obtained for the rejected applicants. However, if the actual situation is similar to the one illustrated by the whole of Figure 3, the individuals who scored relatively low on the predictor test would have received generally lower ratings had they been hired.

In this case, as well as in the Sample Case, the total applicant group is the appropriate reference population for determining the validity of the test, since one wishes to know if the test is helping to make proper decisions about *applicants*. Yet, correlating test scores with criterion ratings for the selected group cannot provide the information about unhired applicants. The test user must realize that *the test may have already done its work*—i.e., many below-average workers have been weeded out by the test and never hired. The range of scores on the predictor test was restricted when only those who scored at the upper end of the score distribution were selected; this led to lower correlation coefficients for the selected group than would have been found if applicants had been hired regardless of their scores.

This kind of restriction resulting from selection is quite common. It occurs not only in industrial settings, but in applied educational research. For example, an investigator who correlates *CEEB Scholastic Aptitude Test* scores with freshman grades for students at a highly selective university may face a problem similar to that of the personnel officer in the Sample Case. The obtained correlation is virtually certain to be much lower than it would be if all of the applicants to the university were admitted.

Q. I see the problem. What can be done to correct for restriction of range?

A. Unfortunately, it is easier to explain the phenomenon of restriction of range than to do something about it. An acceptable, although not always practical, way for a company to study the problem is to administer a test or test battery to all applicants for a limited period of time without using test scores for selection. Instead, for the tryout period, the company either (1) hires all applicants, or (2) hires applicants in virtually a random fashion. Each of these techniques is an attempt to insure that the employees in the validation study will not be restricted in range on the test being evaluated. After an appropriate time interval, scores on the test may be correlated with a criterion such as supervisory ratings or performance in a special training program. (Such correlations are appropriate for estimating the predictive validity of the test, *provided that* the criterion is reasonably reliable.)

Some large organizations have carried out studies similar to the ones suggested above, feeling that the cost of hiring some unproductive employees will be offset by the benefits to be derived from the development of an effective testing program for future selection. However, for other organizations, particularly smaller ones, this kind of procedure is impractical.

During World War II, a boom in testing coincided with a pressing need to select the best-qualified men for specialized wartime positions. As a result, psychologists in the Armed Forces carried out many studies to evaluate the effectiveness of various tests and test batteries.

One frequently-cited wartime study—a project that was carried out as part of the Army Air Force Aviation Psychology Program—illustrates quite clearly the influence of restriction of range (Thorndike, 1949, pp. 170-171). A battery of tests to predict success in pilot training was given to a large group of men. Using the strict selection standards in effect toward the end of the war, only 13 percent of these men would have qualified to enter pilot training on the basis of test scores. Nevertheless, all of the men, regardless of test performance, were allowed to enter training for experimental purposes. Subsequently, correlation coefficients between test scores and the criterion of passing or failing training were computed both for the entire group that had entered training and for the small group that was deemed qualified on the basis of the test scores. The correlation coefficients are shown below for the Composite Aptitude Score and for some of the tests in the battery:

Predictor	Correlation with Criterion*	
	Total Group (N=1036)	Qualified Group (N=136)
Composite Aptitude Score	.64	.18
Mechanical Principles Test	.44	.03
Complex Coordination Test	.40	−.03
Instrument Comprehension Test	.45	.27
Finger Dexterity Test	.18	.00

*Passing or failing training.

These results show the dramatic differences in correlations that can occur when restricted and unrestricted groups are compared. Clearly, if only the qualified group had been studied, the value of the tests would have been grossly underestimated. It may also be noted that even the order of the effectiveness of the predictors was different for the two groups. Judging by the qualified group alone, the *Instrument Comprehension Test* would have seemed to be the best predictor and the *Complex Coordination Test* among the worst; for the total group, however, the *Complex Coordination Test* clearly *was* useful, and the Composite Aptitude Score was the most useful of all.

Q. What can an organization do to determine the effectiveness of a predictor if it cannot carry out a lengthy study?

A. Without a study, there is no way to determine precisely what a validity coefficient would have been had there been no restriction of range resulting from selection on the basis of test scores. It is possible, however, to *estimate* the magnitude of this correlation (R) by a formula which makes use of the actual correlation coefficient obtained from the restricted group (r), and the relationship between the spread of scores of the restricted group and of the total applicant group. Table 1 was developed by using this formula (which appears in a footnote to the table), and facilitates the determination of R_{12}, the estimated correlation between predictor and criterion in a nonrestricted sample.

In the Sample Case, the personnel officer obtained a validity coefficient of .21 (r_{12} in Table 1) for his restricted sample. Since he gave the predictor test to all applicants for the job, he can compute the standard deviation for this group (S_1); he can also compute the standard deviation for the restricted group—the group he hired (s_1). The

TABLE 1
Validity Coefficients for Unrestricted Sample (R_{12})
Estimated from Values for Restricted Sample

$\frac{S_1}{s_1}$ ↓ \ r_{12} →	.10	.15	.20	.25	.30	.35	.40	.45
	Values of R_{12}							
1.25	.12	.19	.25	.31	.37	.42	.48	.53
1.50	.15	.22	.29	.36	.43	.49	.55	.60
1.75	.17	.26	.34	.41	.48	.55	.61	.66
2.00	.20	.29	.38	.46	.53	.60	.66	.71
2.50	.24	.35	.45	.54	.62	.68	.74	.78
3.00	.29	.41	.52	.61	.69	.75	.79	.83
4.00	.37	.52	.63	.72	.78	.83	.87	.90
5.00	.45	.60	.71	.79	.84	.88	.91	.93
10.00	.71	.83	.90	.93	.95	.97	.97	.98

NOTE.—In this table, the symbols $\frac{S_1}{s_1}$, r_{12}, and R_{12} are defined as follows: $\frac{S_1}{s_1}$ is the ratio of the standard deviation of the nonrestricted group (including those selected) to the standard deviation of the restricted group on the predictor test; r_{12} is the actual obtained validity coefficient for the restricted group; R_{12} is the estimated validity coefficient for the total, nonrestricted sample.

The table is read as follows: (1) Locate the value of $\frac{S_1}{s_1}$ closest to the obtained value in the left-hand column. (2) Locate the value of r_{12} closest to the obtained validity in the top row of the table. (3) The point of intersection of the row containing the appropriate $\frac{S_1}{s_1}$ and the column containing the appropriate r_{12} provides R_{12}, the estimated validity coefficient for the nonrestricted group.

For example, if the ratio of standard deviations was found to be 1.90, and if the validity coefficient for the restricted sample was .28, then one should enter the table with the following values: $\frac{S_1}{s_1} = 2.00$ and $r_{12} = .30$. When this is done, $R_{12} = .53$ is read from the intersection of the appropriate row and column for the validity coefficient of the nonrestricted sample.

The values of R_{12} in Table 1 were obtained by the formula (Thorndike, 1949, p. 173):

$$R_{12} = \frac{r_{12}\frac{S_1}{s_1}}{\sqrt{1 - r_{12}^2 + r_{12}^2\frac{S_1^2}{s_1^2}}}$$

ratio between the two standard deviations $\left(\frac{S_1}{s_1}\right)$ may then be computed. If Distribution A of Figure 1 represents the total applicant group, and Distribution B signifies those accepted for work, $\frac{S_1}{s_1}$ is found to be 2.48. Entering the table with values $r_{12} = .20$ and $\frac{S_1}{s_1} = 2.50$ (the tabled values closest to the actual ones), R_{12} is estimated to be .45—substantially higher than the .21 obtained for the restricted group.

Not all possible values of r_{12} and $\frac{S_1}{s_1}$ are presented in Table 1. Users who obtain values that are not specifically shown in the table can use the tabled values that are closest to the ones they obtain (as in the example above), or they can compute R_{12} from the formula.

It should be remembered that R_{12} as thus determined is an *estimate* of the correlation for the total widespread group and that the *true* correlation might be higher or lower if it could actually be obtained. For example, negative or zero values of r_{12} cannot result in positive values of R_{12} due to the nature of the formula.

If the correction formula does not result in a substantial increase in a validity coefficient, it is possible that the test is simply not effective. The test user should consider carrying out a study along the lines discussed earlier or perhaps selecting a new predictor test.

Q. Can the formula be used to correct for low correlations whenever restriction of range occurs?

A. This is a good question, and the answer is an emphatic NO. It is important to remember that the formula (or the table) should be used to provide an estimate of validity *only when a group of people have been selected on the basis of their scores on the predictor test or tests.*

In certain situations, other formulas must be used to correct for restriction of range. For example, suppose that applicants to a college are selected on the basis of scores on Test A, and then a validity study is carried out to see how well Test B (similar in content to Test A) predicts achievement for the selected group. In this instance involving three variables, a formula other than the one discussed here needs to be applied. The interested reader might consult Thorndike (1949, pp. 172-176) or Guilford (1965, pp. 341-345) for a more detailed discussion of this and other special cases.[1]

There are times when it is completely inappropriate to use any formula to correct for restriction of range. As stated above, a formula should *not* be used to correct validity coefficients unless the restricted group in question was *specifically selected from a larger group on a nonrandom basis.*

For example, if a general ability test requiring good reading skills is given to a group of unskilled workers of limited education, it is probable that their scores would fall at the lower end of the possible range. Since the scores of the entire group would be restricted in range, correlations involving the test will be lower for this group than for a group of more widespread ability (e.g., one consisting of unskilled, semi-skilled, and skilled workers). Yet, since scores for the *whole group* of unskilled workers (not just for a selected portion of the group) are involved in the correlation, it is inappropriate and meaningless to use any formula that estimates what the correlation would have been if a more heterogenous group had been measured on both variables.

In a situation such as this, one must conclude that the predictor test used with the group of unskilled workers was much too difficult for them. The director of testing should consider the use of an easier test (one that might spread the raw scores over a wider range) presuming, of course, that his purpose is to distinguish among the group members rather than to diagnose specific problems. In another situation—with a group of gifted individuals, for example—a test may result in range restriction at the upper end of the distribution. The general recommendation is the same as for the unskilled workers: A test at a more appropriate difficulty level is needed.

* * *

The preceding imaginary dialogue was intended to clarify a topic that is a thorny problem for applied researchers. Remember, however, that although restriction of range can be quite influential, it is only one of many factors than can depress correlation coefficients. As stated previously, the two variables being correlated may truly have nothing in common, regardless of the heterogeneity of the group. For validity coefficients, it is possible that the criterion being used is not reliable; this would reduce correlations with the predictor test for any group of individuals.

Clearly, restriction of range is important and test users should be aware of what it is and what its effects are. When a test is used for selection, as in the Sample Case, then restriction of range is certain to enter into the picture. When tests are used for other purposes, one should always take the time to tabulate the scores for any sizable group in order to observe the distribution pattern and note any restriction of obtained scores. A convenient standard for judging "full range" might be a group of applicants who apply, or a norms group reported in the test manual. When the range of scores is restricted, it is important to note the specific pattern. If most of those tested obtain very high scores, the test may be too easy for the group (though it might be useful if the purpose is only to identify a small number of people for remedial help or for exclusion from the job). If the scores are clustered at the low end of the distribution, the test is probably too hard for this group unless the examiner wishes only to identify the relatively small group of high scorers. In all cases, prudence requires us to take the effect of restriction of range into account in making our decisions.

[1] For a technical discussion of an interesting paradox in which, under certain circumstances, the corrected coefficients in three-variable cases may actually be *lower* than the obtained correlations, see Levin (1972).

REFERENCES

Guilford, J. P. *Fundamental statistics in psychology and education.* (4th ed.) New York: McGraw-Hill, 1965.

Levin, J. The occurrence of an increase in correlation by restriction of range. *Psychometrika,* 1972, **37,** 93-97.

Thorndike, R. L. *Personnel selection: Test and measurement techniques.* New York: Wiley, 1949.

The author gratefully acknowledges the help of Mr. Louis B. Jones in the discussions and drafts leading to this article.

A New Manual with Useful Data for the

Ψ SHORT EMPLOYMENT TESTS

GEORGE K. BENNETT and MARJORIE GELINK

Research was the foundation of the *Short Employment Tests*–specifically, a series of studies of the selection of clerical employees conducted for the member banks of the American Bankers Association. Research reported in the first and second editions of the *SET* manual supported the growth that has made *SET* reportedly the battery of tests most widely used in the selection of office workers.

Continuing research over a sixteen-year period has provided the data now presented in the new 1972 revision of the manual for the *SET*. Carried out chiefly by conscientious users of the tests, the studies increase to 56 the number of norm groups (applicants and employees) with which scores can be compared, and to 63 the illustrative validation studies. The results of these studies will, we hope, encourage other users to undertake their own. Data for minority groups are reported separately in the new manual wherever possible.

The *SET* combine in three five-minute tests–Verbal, Numerical, and Clerical Aptitude–the kinds of scores that are the most effective predictors of satisfactory performance in office work. Norms and validity data are given for each of the three separate tests as well as for the total score.

The new manual is being supplied with all shipments of the *SET*. Users already having a supply of the tests may obtain a copy free by requesting it. A Specimen Set of the *SET* (**2B811**) is **$1.00**.

---

and a Newly Revised Manual also for the

Ψ GENERAL CLERICAL TEST

More elaborate and lengthier than the *Short Employment Tests*, above, the *General Clerical Test* affords a correspondingly more thorough appraisal of the aptitudes and abilities important in various kinds of office work. Like the *SET*, the *GCT* now has a new manual which greatly enriches the store of data on which users may base their judgments of the test's applicability and their interpretations of the meaning of scores.

Eleven new norm groups have been added to the 30 available in the preceding edition. Users of the *GCT* in most employment situations and many academic situations will be able to find among these groups one or more with which their examinees may fairly be compared. The validation data reported have been similarly increased, including a two-page table presenting results of sixteen on-the-job studies, and another table presenting score averages for groups engaged in jobs of contrasting difficulty and responsibility.

The new *GCT* manual is currently sent with all orders for the test, or free on request to any user. A Specimen Set of the *General Clerical Test* (**2A805**) is **90** cents.

HANDBOOK for the STRONG VOCATIONAL INTEREST BLANK

DAVID P. CAMPBELL, University of Minnesota

In his Preface to this 544-page *Handbook*, the author comments, "This . . . is a succession of dreary statistics about the most exciting topic in the world: a man's enthusiasm for his work. . . . The excitement of occupation has led men into more risks . . . than could money or fame; . . . the possibility of better employment has created larger mass migrations than has religious fervor, and the absence of meaningful work has likely created more mental depression than any other single factor. For most people, where they work is where the action is."

Not all so dreary, though. Several thousand means, standard deviations, and correlations are indeed reported in the 375 tables and figures, and these data are of fundamental importance to those who use interest inventories, particularly (though not only) the *Strong*. There is much, much more than numbers, however, in the eleven chapters and six appendices.

Individual case studies of men and women, discussions of the interests of outstanding and special groups, and keen observations on the measurement problems of appraising vocational interests supplement the descriptions of the occupational and other scales. The stability and change of interests over time, the possibility of faking scores, and the counseling interpretation of *SVIB* profiles all receive due attention. Going far beyond the *SVIB Manual* and *Supplement*, the *Handbook for the Strong Vocational Interest Blank* is a necessary and a rewarding work for both counselor and researcher. Order: **9V748 $20.00** plus postage

---

Revised and Restandardized Third Edition of the

COLUMBIA MENTAL MATURITY SCALE

B. B. BURGEMEISTER, L. H. BLUM, and I. LORGE, Teachers College, Columbia University

Thoroughly revised and restandardized, the Third (1972) Edition of the *Columbia Mental Maturity Scale* is available for use with young children who, by reason of neurological impairment or other handicap, cannot respond to ordinary intelligence test questions well enough to obtain a meaningful score. The *CMMS* requires no verbal response and only a minimum motor response.

Since no reading is required and since the response may be either verbal or by gesture, the *CMMS* permits the testing of children who have cerebral palsy, brain damage, or impairments of speech and hearing—or who simply come from backgrounds where English is not spoken or where other educational experience has been inadequate. Directions in Spanish are included in the manual, and the visual nature of the test is such that it can be given with the aid of an interpreter in other languages; it is not subject to the disadvantages of foreign language administration of tests dependent on precise oral instructions and repeated oral questions.

Although specialized in nature and purpose, the *CMMS* has been standardized on normal children aged 3½ to 10. The norms are based on a national sample selected to be representative of parental occupation, race, and geographic and urban/rural residence. The score may be reported either as a percentile within age group or as a Maturity Index (age equivalent).

The ninety-five cards constituting the test materials have been classified in eight overlapping age or ability levels so that most children need be given only that portion most appropriate to their capability. Testing ordinarily requires no more than twenty minutes.

8H017	Examiner's Kit, complete	**$50.00**
8H902	Kit, as above, in carrying case	**$62.00**

Alternate Form of the Boehm

The new Form B of the *Boehm Test of Basic Concepts* measures knowledge of the same concepts as Form A; both item responses and the total score may be compared as indices of progress in learning. A Specimen Set (**5L806**) is **$1.00.**

Two New Books for Users of the

MINNESOTA MULTIPHASIC PERSONALITY INVENTORY

In the world of tests and inventories, the *Minnesota Multiphasic Personality Inventory* has been unique in a number of ways. Not the least of these is the fact that the basic information and interpretive material supporting its use could not be compressed into an ordinary test manual; it occupies a collection of books now numbering about a dozen.

Among these dozen books, the one entitled *An MMPI Handbook*, by W. Grant Dahlstrom and George S. Welsh, has occupied a central position. Subtitled "A Guide to Use in Clinical Practice and Research," it has since its appearance in 1960 generally been referred to in MMPI circles as simply "**The** Handbook." Now in 1972, it is being superseded by a greatly enriched and expanded two-volume revised edition, of which Volume I is now in stock.

Although subtitled "Clinical Interpretation" and devoting primary attention to clinical use of the MMPI, (as distinguished from research applications, to be the focus of Volume II), Volume I includes the important tables and appendices that have made the earlier edition invaluable. Of all MMPI books, this is the one a serious user can least afford to be without. Order: **9V255 $19.00** plus postage

Equally rich, if less absolutely essential, is the material presented by 21 authors in *MMPI: Research Developments and Clinical Applications*, edited by James N. Butcher. From technical developments such as computer-based interpretation to social issues such as the question of invasion of privacy, this book provides a stimulating sample of work and thought. It includes a useful 756-item classified bibliography. Order: **9V085 $11.20** plus postage

Indispensable . . .

THE SEVENTH MENTAL MEASUREMENTS YEARBOOK

OSCAR K. BUROS, EDITOR

"Indispensable"—the word in the headline above the title—is without much doubt the adjective on which it would be easiest to obtain agreement among competent makers and users of tests with regard to the series of *Mental Measurements Yearbooks*. The new *Seventh* comprises all the values of its predecessors, but in larger measure than any of them.

The more than two thousand pages are bound, fortunately, in two volumes. The simple, clear indexing system gives the possessor of the *Seventh* ready access to the 6,400-plus pages of the first six editions.

As those who have seen any of the earlier *Yearbooks* know, statistics can give only the most inadequate picture. Yet it would be wrong to withhold the information that the newest edition lists 1,157 tests with 12,372 references. For 546 of the tests, there are 798 original reviews by 439 reviewers, supplemented by another 181 reviews excerpted from 39 journals.

The diversity of the reviewers and the frankly critical nature of most of the reviews is a mark of this *Yearbook* as of those that have gone before. The editor's experience and high standards are nowhere more evident than in the frequency with which the most important and widely used tests have been assigned to the most experienced reviewers. The editor's equally high standards of bookmaking and of accuracy in detail are legendary. Innovations in *The Seventh MMY* include reviews of certain scoring machines and scoring/reporting services in addition to the tests themselves.

A special case of professional illiteracy is the fate of any teacher or student of educational and psychological tests and measurements who does not have access to Dr. Buros's *Seventh Mental Measurements Yearbook*. Any psychometrician or psychometrist who cannot afford it is urged to talk—promptly and strongly—with his librarian.
Order: **9V762** **$57.00** plus postage

Good Minds at Work on a Difficult and Timely Problem

INTELLIGENCE: GENETIC AND ENVIRONMENTAL INFLUENCES

ROBERT CANCRO, Editor, University of Connecticut School of Medicine

A remarkable group of people assembled on the campus of the University of Illinois for a conference on intelligence. This volume presents the principal papers prepared for the conference.

The heritability of intelligence, the effects of environment and of socioeconomic forces, and the nature and extent of differences in intelligence between the sexes and among the races are the topics upon which the meeting focussed. Fifteen papers are grouped in three categories: Theory and Measurement, Genetic Contributions, and Environmental Contributions.

The depth and quality of the conference can best be suggested by listing the authors of these papers, including David Wechsler, Arthur Jensen, J. McV. Hunt, C. C. Li, Lloyd Humphreys, and R. B. Cattell, as well as Sidney Bijou, Robert Cancro, Bruce Eckland, Benson Ginsburg, Edmund Gordon, Patricia Greenfield, Jerry Hirsch, Girvin Kirk, William Laughlin, Philip Merrifield, and Steven Vandenberg.

The argument over "nature vs. nurture" is not new, but has never been more central to the liveliest educational and social issues of the day. Calm and informative rather than argumentative, these papers bring the light of science to bear in an effort to either overcome or clarify some of the common misunderstandings. Titles among the fifteen include:

Race and Intelligence: What Do We Really Know?
The Race X Sex X Ability Interaction, and
What Do We Know Today About the Inheritance of Intelligence and How Do We Know It?

Interesting and provocative, scientifically sound, the book offers good insurance against the risk of falling behind the advance of thought in a most important area.
Order: **9V798** **$12.75** plus postage

Spanish Editions Include a New Translation of the

Ψ SURVEY OF STUDY HABITS AND ATTITUDES

W. F. BROWN and W. H. HOLTZMAN, University of Texas

A Spanish version of the Brown-Holtzman *SSHA*, the *Encuesta de Hábitos y Actitudes hacia el Estudio (EHAE)*, has been prepared by Fernando García Cortés and Eduardo García Hassey, with a manual translated and adapted by Luis M. Laosa. It is the latest addition to the growing number of tests listed in THE PSYCHOLOGICAL CORPORATION's Test Catalog for use with persons more at home in Spanish than in English.

Using students from a number of Hispano-American countries in the experimental tryouts, the authors and translators have taken pains to avoid localisms or nationalisms so as to make the *EHAE* readily applicable in most Central and South American nations as well as with Chicanos and Puerto Ricans in the United States. A single form serves both high school students and college freshmen. A Specimen Set (**4G813**) is **$1.35**.

Others in the list of tests available in Spanish as well as English appearing on page 58 of the 1972 Test Catalog include:

Differential Aptitude Tests
Bennett Mechanical Comprehension Test
Barranquilla Rapid Survey Intelligence Test*
Boehm Test of Basic Concepts
PTI-Oral Directions Test
Wechsler Adult Intelligence Scale (EIWA)
Wechsler Intelligence Scale for Children (EIWN)
Minnesota Multiphasic Personality Inventory

*BARSIT is available *only* in Spanish

PUPIL RATING SCALE FOR LEARNING DISABILITIES

H. R. MYKLEBUST, Northwestern University

This five-point rating scale is designed to assist the classroom teacher in detecting learning disabilities in children 7 through 10 years of age. Careful definitions of behavioral characteristics help teachers make and record accurate observations in the areas of Auditory Comprehension, Spoken Language, Orientation, Motor Coordination, and Personal-Social Behavior.

8J404	Manual (with copy of Record Form)	**$4.20**
8J208	Record Forms, package of 50	**$2.60**

A Handbook for Clinical Application

THE HOLTZMAN INKBLOT TECHNIQUE

EVELYN F. HILL, Johns Hopkins University

"The *Holtzman Inkblots* . . . [were] introduced only ten years ago, and have such obvious advantages over other projective inkblot tests that [the *HIT*] is already rapidly replacing the older techniques, particularly in the field of research. The ever-expanding *HIT* literature also attests to its impact on clinical psychology. Those using the *Holtzman* have long felt the need for a complete guide to the interpretation of *HIT* scores. . . . after using the *HIT* for more than ten years, both in the clinic and in research, we are now able to present material which gives the clinical psychologist a highly sensitive projective tool." Sometimes the author's own prefatory remarks give both the briefest and the best summary of a work; an instance of this appears in the quoted words above from Dr. Hill's preface.

The Holtzman Inkblot Technique is a real "handbook" in the sense in which Dahlstrom and Welsh have made the term meaningful for the *MMPI*. It is at the same time a general guide, a clinical manual, and a teaching text. Since the original monograph by Holtzman and his colleagues appeared before the possibilities of using *HIT* scores in assessing personality had been fully explored, this book provides the only really comprehensive interpretation of *HIT* scores. Order: **9V047** **$17.75** plus postage

The contents of the handbook are divided into four parts: I—Background and Technique, II—Interpretation Based on Scoring Variables, III—Interpretation Based on Personality Variables, and IV—Case Analyses. An appendix presents normative data for fifteen sample populations (seven normal, eight abnormal). A special form, the *Hill Clinical Summary for the Holtzman Inkblot Technique*, is to be published separately for regular use by psychologists.

A New Six-Score Individual Examination for Young Children

Ψ McCARTHY SCALES OF CHILDREN'S ABILITIES

DOROTHEA A. MCCARTHY

A more comprehensive measure of the cognitive and behavioral development of young children than has been available from any other one instrument was the aim of the new *McCarthy Scales of Children's Abilities*. For children aged 2½ to 8½, the *MSCA* provides scores or Indexes for these six Scales:

VERBAL—the ability to deal with concepts in words and to express ideas

PERCEPTUAL-PERFORMANCE—the ability to reason with concrete materials

QUANTITATIVE—the ability to use numbers, in counting and in solving problems

GENERAL COGNITIVE—an index of mental competence and academic readiness based on verbal, perceptual-performance, and quantitative tasks

MEMORY—short-term, both visual and auditory

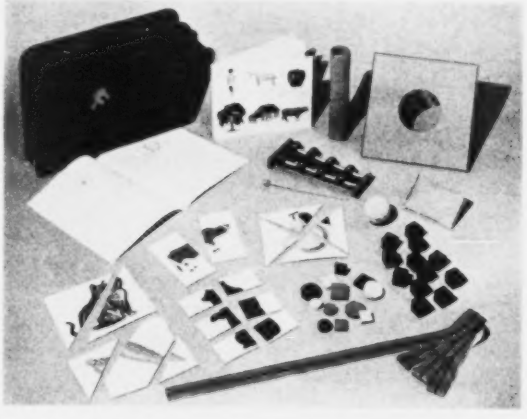
MOTOR—gross and fine coordination, and opportunities to observe lateral dominance.

Dr. McCarthy, professor emeritus at Fordham University and former director of its Child Guidance Clinic, has designed eighteen brief subtests that hold the child's attention; the materials and questions are game-like and non-threatening. Items sample a wide range of behavior so that children proceed easily from one to another with little risk of overtaxing a short attention span. The entire examination, individually administered, usually takes no more than an hour.

With the *McCarthy*, how well a child is likely to do in school and in life in the years immediately ahead can be estimated on the basis of his present functioning. The six Scales afford both a well-rounded understanding of the child and a foundation for differential diagnosis of possible neurological impairment or learning disorders.

Norms for the *McCarthy Scales* have been established by testing a nationwide sample of children. The sample, based on 1970 Census data, was stratified on the variables of age, sex, color, geographic region, and father's occupation. Minorities sampled included children from Spanish-speaking families as well as blacks and other nonwhite groups. In addition, children were selected from urban and rural areas in approximate proportion to the Census figures.

Dr. McCarthy's experience has enabled her to prepare a manual that offers, in effect, a short course in the administration of tests to young children, particularly to preschool children who may have had no previous experience with tests. The manual also includes, of course, complete directions for each task and the necessary norms and other data. The author and the staff have given special attention to designing the record form in such a way as to ease the work of the examiner.



Dr. McCarthy has avoided use of the term Intelligence Quotient or "IQ" in the provisions for reporting results on the *MSCA*. Instead there is a General Cognitive Index which has a mean of 100 and a standard deviation of 16 for convenience of interpretation. The Index for each of the other five Scales has a mean of 50 and a standard deviation of 10. Supplementary observations, such as of right- or left-handedness, are recorded separately from the numerical scores.

It is expected that complete test kits, including carrying case, the manual, and a supply of record forms, will be ready for shipment in June.

6D019	Complete Set	**$58.00**	
6D409	Manual, separately	**$5.00**	
6D215	Record Forms, package of 25		**$3.50**
6D227	Package of 100	**$12.00**	
6D071	Drawing Booklets, package of 25		**$4.75**
6D083	Package of 100	**$15.50**	